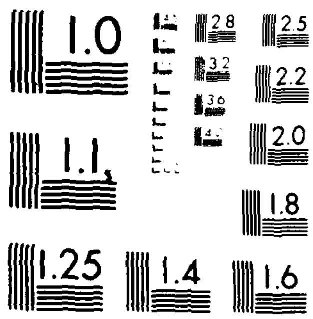
AD-A131 804 ARTIFICIAL INTELLIGENCE: COGNITION AS COMPUTATION(U) 1/1
STANFORD UNIV CA DEPT OF COMPUTER SCIENCE A BARR
AUG 82 STAN-CS-82-956 N00014-79-C-0302
UNCLASSIFIED F/G 9/2 NL


1.0
1.1
1.25
1.4
1.6
1.8
2.0
2.2
2.5
2.8
3.2
3.6
4.0

August 1982
Also numbered HPP-82-29

Report No. STAN-CS-82-956

Artificial Intelligence: Cognition as Computation

by

Avron Barr

Department of Computer Science

Stanford University
Stanford, CA 94305

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER See face of report	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) See report		5. TYPE OF REPORT & PERIOD COVERED See report
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) See report and Work unit		8. CONTRACT OR GRANT NUMBER(s) See part 17 of work unit
9. PERFORMING ORGANIZATION NAME AND ADDRESS N W Coordinate with Work Unit		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS See part 10 of work unit
11. CONTROLLING OFFICE NAME AND ADDRESS ONR Code in part ~~20~~ 1C of work unit		12. REPORT DATE See report
		13. NUMBER OF PAGES
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) UNCLASSIFIED
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

This report was generated with the use of government funds.

Work unit is attached.

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

See report and work unit.

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

See report and work unit.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-014-6601

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

```
--    1 - AGENCY ACCESSION NO:    DN975390
--   12 - SUMMARY RECEIPT DATE:   30 JUN  82
--    2 - DATE OF SUMMARY: 28 JUN  82
--    3 - DATE OF PREV. SUMMARY: 01 NOV  79
--    4 - KIND OF SUMMARY: TERMINATED
--    5 - SUMMARY SECURITY IS:   UNCLASSIFIED
--    6 - SECURITY OF WORK:      UNCLASSIFIED
--  8A1 - DIST. LIMITATION: UNLIMITED
--   8B - CONTRACTOR ACCESS:     YES
--10A1 - PRIMARY PROGRAM ELEMENT:      62709E
--10A2 - PRIMARY PROJECT NUMBER:       9W10
--10A2A - PRIMARY PROJECT AGENCY AND PROGRAM:      9W10
--10A3 - PRIMARY TASK AREA:      3697
--10A4 - WORK UNIT NUMBER:       NR-154-436
--10B1 - CONTRIBUTING PROGRAM ELEMENT (1ST):      61153N
--10B2 - CONTRIBUTING PROJECT NUMBER (1ST): RR04206
--10B3 - CONTRIBUTING TASK AREA (1ST):       RR0420601
--   11 - TITLE:      (U)  PERSONNEL TECHNOLOGY: TUTORING AND PROBLEM-SOLVING
--      STRATEGIES IN INTELLIGENT COMPUTER-AIDED INSTRUCTION
-- 11A - TITLE SECURITY: U
--   12 - S + T AREAS:
--                  012500 PERS SELECTION,TRAINING & EVAL

--

--                  013400 PSYCHOLOGY-INDIV/GROUP BEHAVIOR
--   13 - WORK UNIT START DATE:  MAR 79
--   14 - ESTIMATED COMPLETION DATE:       CD
-- 15A - PRIMARY FUNDING AGENCY:  DEPT. OF DEFENSE
-- 15B - OTHER FUNDING AGENCY:  NAVY
--   16 - PERFORMANCE METHOD:    CONTRACT
--17A1 - CONTRACT/GRANT EFFECTIVE DATE:      MAR 79
--17A2 - CONTRACT/GRANT EXPIRATION DATE:     SEP 80
-- 17B - CONTRACT/GRANT NUMBER: N00014-79-C-0302
-- 17C - CONTRACT TYPE:    COST TYPE
--17D2 - CONTRACT/GRANT AMOUNT:      $ 166,000
-- 17E - KIND OF AWARD:    SUP
-- 17F - CONTRACT/GRANT CUMULATIVE DOLLAR TOTAL:      $ 75,447
--                 RESOURCE ESTIMATES
--18A4 MANYRS CFY+1:        +   18B4 FUNDS CFY+1:          +
--18A3 MANYRS CFY-3:      2.6   18B3 FUNDS CFY-3:  $ 1,406,000
--18A2 MANYRS CFY-2:      0.3   18B2 FUNDS CFY-2:    $ 166,000
--18A1 MANYRS CFY-1:        +   18B1 FUNDS CFY-1:          +
--18A  MANYRS CFY:          +   18B  FUNDS CFY:            +
-- 19A - DOD ORGANIZATION:      OFFICE OF NAVAL RESEARCH (458)
-- 19B - DOD ORG. ADDRESS:      ARLINGTON, VA 22217
-- 19C - RESPONSIBLE INDIVIDUAL:      FARR, M.J.
-- 19D - RESPONSIBLE INDIVIDUAL PHONE:   202-696-4504

--

-- 19U - DOD ORGANIZATION LOCATION CODE:  5110
-- 19S - DOD ORGANIZATION SORT CODE:   35832
-- 19T - DOD ORGANIZATION CODE: 265250
```

DTIC COPY INSPECTED

Accession For
NTIS GRA&I X
DTIC TAB
Unannounced
Justification

By
Distribution/
Availability Codes

Dist	Avail and/or Special
A	

-- 20A - PERFORMING ORGANIZATION: STANFORD UNIVERSITY COMPUTER SCIENCE
-- DEPT
-- 20B - PERFORMING ORG. ADDRESS: STANFORD, CA 94305
-- 20C - PRINCIPAL INVESTIGATOR: BUCHANAN, B G
-- 20D - PRINCIPAL INVESTIGATOR PHONE: 415-497-0935
-- 20F - ASSOCIATE INVESTIGATOR (1ST): CLANCEY, W
-- 20G - ASSOCIATE INVESTIGATOR (2ND): BARR, A
-- 20U - PERFORMING ORGANIZATION LOCATION CODE: 0612
-- 20N - PERF. ORGANIZATION TYPE CODE: 1
-- 20S - PERFORMING ORG. SORT CODE: 44744
-- 20T - PERFORMING ORGANIZATION CODE: 094120
-- 22 - KEYWORDS: (U) PRODUCTION SYSTEM ;(U) COMPUTER-ASSISTED INSTRUCTION ;
-- (U) CAI ;(U) ARTIFICIAL INTELLIGENCE ;(U) PROBLEM SOLVING ;(U)
-- MEDICAL DIAGNOSIS ;(U) MEDICAL TRAINING ;(U) MYCIN ;
-- 23 - TECHNICAL OBJECTIVE: (U) MANY NAVY AND MARINE CORPS JOBS (E.G.,
-- ELECTRONIC TROUBLESHOOTING) REQUIRE HIGHLY DEVELOPED DIAGNOSTIC AND
-- PROBLEM-SOLVING SKILLS. THIS WORK HAS SEVERAL PRIMARY OBJECTIVES: (1) TO
-- INVESTIGATE DOMAIN-INDEPENDENT METHODS FOR TEACHING STRATEGIES FOR
-- DEVISE FORMAL, COMPUTATIONAL MODELS FOR PLANNING AND EXECUTING TUTORIAL
-- DIALOGS; (3) TO EVALUATE THE EFFECTIVENESS OF DIFFERENT MODES OF

-

-- TUTORIAL INTERACTION; AND (4) TO FORMULATE FORMAL MODELS FOR HUMAN
-- LEARNING OF RULE-BASED KNOWLEDGE.
-- 24 - APPROACH: (U) AN INITIAL VERSION OF AN INSTRUCTIONAL SYSTEM FOR
-- TEACHING THE DIAGNOSIS OF BACTERIAL INFECTIONS WILL BE TESTED WITH
-- STUDENTS OF VARYING ABILITY, AND THE RESULTS USED TO EXTEND AND REFINE A
-- SET OF DOMAIN-INDEPENDENT TEACHING RULES. EXPERIMENTS WILL BE CONDUCTED
-- TO EVALUATE THE EFFECTIVENESS OF ALTERNATIVE TUTORIAL STRATEGIES.
-- ADDITIONAL CAPABILITIES FOR PLANNING AND EXECUTING INSTRUCTIONAL DIALOGS,
-- WITH SPECIFIC ATTENTION TO INDIVIDUALIZED INSTRUCTION, WILL BE
-- INCORPORATED INTO THE INSTRUCTIONAL SYSTEM. ONE OR MORE ADDITIONAL
-- SYSTEMS FOR TEACHING DIAGNOSIS IN THE CONTEXT OF ELECTRONIC OR
-- MECHANICAL TROUBLESHOOTING PROBLEMS WILL BE DEVELOPED AND EXPERIMENTED
-- WITH TO EVALUATE THE GENERALITY OF THE KERNEL INSTRUCTIONAL SYSTEM.
-- 25 - PROGRESS: (U) A DOMAIN-INDEPENDENT MODEL OF DIAGNOSTIC REASONING WAS
-- FORMALIZED THAT THE GUIDON TUTORIAL MONITOR WILL EXPLICITLY CONVEY TO
-- STUDENTS. THE TOP-LEVEL PLAN EMPHASIZES COMPLETENESS AND CAREFUL
-- REFINEMENT OF THE HYPOTHESIS SPACE. GUIDON'S MODEL OF THE STUDENT HAS
-- BEEN REDESIGNED TO DETECT USAGE PATTERNS OF THE STUDENT, AS THE
-- FIRSTSTEP IN CONSTRUCTING A MODEL OF HIS PROBLEM-SOLVING, EXPLORATIVE
-- PROBLEM-SOLVING, KNOWLEDGE-LIMITED PROBLEM-SOLVING, AND STRATEGICAL
-- PROBLEM-SOLVING.
-- 30 - SUBELEMENT CODE: 42
-- 37 - DESCRIPTORS: (U) BACTERIAL DISEASES;(U) ARTIFICIAL INTELLIGENCE ;

-

-- (U) TRAINING ;(U) TEACHING METHODS ;(U) STUDENTS ;(U) SKILLS;
-- (U) PRODUCTION ;(U) PROBLEM SOLVING ;(U) PERSONNEL ;(U)
-- MEDICINE ;(U) MATHEMATICAL MODELS ;(U) LEARNING ;(U)
-- INDIVIDUALIZED TRAINING ;(U) DIAGNOSIS(MEDICINE) ;(U)
-- DIAGNOSIS(GENERAL);(U) COMPUTER AIDED INSTRUCTION ;(U) COMPUTATIONS;
-- 39 - PROCESSING DATE (RANGE): 30 JUN 82
--******
-- 2 OF 3

Artificial Intelligence: Cognition as Computation[1]

Avron Barr

The ability and compulsion to *know* are as characteristic of our human nature as are our physical posture and our languages. Knowledge and intelligence, as scientific concepts, are used to describe how an organism's experience appears to mediate its behavior. This report discusses the relation between artificial intelligence (AI) research in computer science and the approaches of other disciplines that study the nature of intelligence, cognition, and mind. The state of AI after 25 years of work in the field is reviewed, as are the views of its practitioners about its relation to cognate disciplines. The report concludes with a discussion of some possible effects on our scientific work of emerging commercial applications of AI technology, that is, machines that can know and can take part in human cognitive activities.

Artificial Intelligence

Artificial intelligence is the part of computer science concerned with creating and studying computer programs that exhibit behavioral characteristics we identify as intelligent in human behavior—knowing, reasoning, learning, problem solving, language understanding, and so on. Since the field's emergence in the mid-1950s, AI researchers have developed dozens of programs and programming techniques that support some sort of "intelligent" behavior. Although there are many attitudes expressed by researchers in the field, most of these people are motivated in their work on intelligent computer programs by the thought that this work may lead to a new understanding of mind:

> AI has also embraced the larger scientific goal of constructing an information-processing theory of intelligence. If such a *science of intelligence* could be developed, it could guide the design of intelligent machines as well as explicate intelligent behavior as it occurs in humans and other animals. (Nilsson, 1980, p. 2)

---

[1] To appear in *The Study of Information: Interdisciplinary Messages*, edited by Fritz Machlup and Una Mansfield, and published by John Wiley and Sons, New York, 1983.

Whether or not it leads to a better understanding of the mind, there is every evidence that current work in AI will lead to a new *intelligent technology* that may have dramatic effects on our society. Experimental AI systems have already generated interest and enthusiasm in industry and are being developed commercially. These experimental systems include programs that—

- solve some hard problems in chemistry, biology, geology, engineering, and medicine at human-expert levels of performance;
- manipulate robotic devices to perform some useful sensory-motor tasks; and
- answer questions posed in restricted dialects of English (French, Japanese, etc.).

Useful AI programs will play an important part in the evolution of the role of computers in our lives—a role that has changed, in our lifetimes, from remote to commonplace and that, if current expectations about computing cost and power are correct, is likely to evolve further from useful to essential.

The Origins of Artificial Intelligence

> Scientific fields emerge as the concerns of scientists congeal around various phenomena. Sciences are not defined, they are recognized. (Newell, 1973a, p. 1)

The intellectual currents of the times help direct scientists to the study of certain phenomena. For the evolution of AI, the two most important forces in the intellectual environment of the 1930s and 1940s were *mathematical logic,* which had been under rapid development since the end of the 19th century, and new ideas about *computation.* The logical systems of Frege, Whitehead and Russell, Tarski, and others showed that some aspects of reasoning could be formalized in a relatively simple framework:

> The fundamental contribution was to demonstrate by example that the manipulation of symbols (at least *some* manipulation of *some* symbols) could be described in terms of specific, concrete processes quite as readily as could the manipulation of pine boards in a carpenter shop.... Formal logic, if it showed nothing else, showed that ideas—at least some ideas—could be represented by symbols, and that these symbols could be altered in meaningful ways by precisely defined processes. (Newell and Simon, 1972, p. 877)

Mathematical logic continues to be an active area of investigation in AI, in part because general-purpose, logico-deductive systems have been successfully implemented on computers. But even before the advent of computers, the mathematical formalization of logical reasoning shaped people's conception of the relation between computation and intelligence.

Ideas about the nature of computation, due to Church, Turing, and others, provided the iink between the notion of formalization of reasoning and the computing machines about to be invented. What was essential in this work was the abstract conception of computation as *symbol processing*. The first computers were numerical calculators that did not appear to embody much intelligence at all. But before these machines were even designed, Church and Turing had seen that numbers were an inessential aspect of computation—they were just one way of interpreting the internal states of the machine:

> In their striving to handle symbols rigorously and objectively—as objects—logicians became more and more explicit in describing the processing system that was supposed to manipulate the symbols. In 1936, Alan Turing, an English logician, described the processor, now known as the *Turing machine*, that is regarded as the culmination of this drive toward formalization. (Newell and Simon, 1972, p. 878)

> The model of a Turing machine contains within it the notions both of what can be computed and of universal machines—computers that can do anything that can be done by any machine. (Newell and Simon, 1976, p. 117)

Turing, who has been called the father of AI, not only invented a simple, universal, and nonnumerical model of computation but also argued directly for the possibility that computational mechanisms could behave in a way that would be perceived as intelligent:

> Thought was still wholly intangible and ineffable until modern formal logic interpreted it as the manipulation of formal tokens. And it seemed still to inhabit mainly the heaven of Platonic ideals, or the equally obscure spaces of the human mind, until computers taught us how symbols could be processed by machines. A. M. Turing . . . made his great contributions at the mid-century crossroads of these developments that led from modern logic to the computer. (Newell and Simon, 1976, p. 125)

As Allen Newell and Herbert Simon point out in the "Historical Epilogue" to their classic work *Human Problem Solving* (1972), there were other strong intellectual currents from several directions that converged in the middle of this century in the people who founded the science of artificial intelligence. The concepts of cybernetics and self-organizing systems of Wiener, McCulloch, and others dealt with the macroscopic behavior of "locally simple" systems. The cyberneticians influenced many fields because their thinking spanned many fields, linking ideas about the workings of the nervous system with information theory and control theory, as well as with logic and computation. Their ideas were part of the zeitgeist, but in many cases the cyberneticians influenced early workers in AI more directly—as their teachers.

What eventually connected these diverse ideas was, of course, the development of the computing machines themselves, conceived by Babbage and guided in this century by Turing, von Neumann, and others. It was not long after the machines became available that people began to try to write programs to solve puzzles, play chess, and translate texts from one language to another—the first AI programs.

What was it about computers that triggered the development of AI? Many ideas about computing relevant to AI emerged in the early designs—ideas about memories and processors, about systems and control, and about levels of languages and programs. But the single attribute of the new machines that brought about the emergence of the new science was their inherent potential for *complexity*, encouraging (in several fields) the development of new and more direct ways of describing complex processes—in terms of complicated data structures and procedures with hundreds of different steps:

> Problem solving behaviors, even in the relatively well-structured task environments that we have used in our research, have generally been regarded as highly complex forms of human behavior—so complex that for a whole generation they were usually avoided in the psychological laboratory in favor of behaviors that seemed to be simple. . . . The appearance of the modern computer at the end of World War II gave us and other researchers the courage to return to complex cognitive performances as our source of data . . . a device capable of symbol-manipulating behavior at levels of complexity and generality unprecedented for man-made mechanisms. . . . This was part of the general insight of cybernetics, delayed by ten years and applied to discrete symbolic behavior rather than to continuous feedback systems. (Newell and Simon, 1972, pp. 869-870)

Computers, Complexity, and Intelligence

As Pamela McCorduck notes in her entertaining historical study of AI *Machines Who Think* (1979), there has been a longstanding connection between the idea of complex mechanical devices and intelligence. Starting with the fabulously intricate clocks and mechanical automata of past centuries, people have made an intuitive link between the *complexity* of a machine's operation and some aspects of their own mental life. Over the last few centuries, new technologies have resulted in a dramatic increase in the complexity we can achieve in the things we build. Modern computer systems are more complex by several orders of magnitude than anything humans have built before.

The first work on computers in this century focused on the numerical computations that had previously been performed collaboratively by teams of hundreds of clerks, organized so that each did one small subcalculation and passed the results on to the clerk at the next desk. Not long after the dramatic success of the first digital computers with these elaborate calculations, people began to explore the possibility of more generally intelligent mechanical behavior—could machines play chess, prove theorems, or translate languages? They could, but not very well. The computer performs its calculations following the step-by-step instructions it is given—the method must be specified *in complete detail*. Most computer scientists are concerned with designing new algorithms, new languages, and new machines for performing tasks like solving equations and alphabetizing lists—tasks that people perform using methods they can explicate. However, people cannot specify how they decide which move to make in a game of chess or how they determine that two sentences "mean the same thing."

The realization that the detailed steps of almost all intelligent human activity were unknown marked the beginning of artificial intelligence as a separate part of computer science. AI researchers investigate different kinds of computation, and different ways of describing computation, in an attempt not just to create intelligent artifacts but also to understand what intelligence is. A basic tenet of AI is that human intellectual capacity will best be described in the same terms as the ones researchers invent to describe their programs. However, they are just beginning to learn enough about those programs to know how to describe them scientifically—in terms of concepts that illuminate their nature and differentiate among fundamental categories. These ideas about computation have been developed in programs that perform many different tasks, sometimes at the level of human performance, often at a much lower level. Most of these methods are obviously not the same as the ones that people use to perform the tasks—some of them might be.

The Status of Artificial Intelligence

Many intelligent activities besides numerical calculation and information retrieval have been carried on by programs. Many key aspects of thought—like recognizing people's faces and reasoning by analogy—are still puzzles; they are performed so unconsciously by people that adequate computational mechanisms have not been postulated. Some of the successes, as well as some of the failures, have come as surprises. We will list here some of the aspects of intelligence investigated in AI research and try to give an indication of the stage of progress.

There is an important philosophical point here that will be sidestepped. Doing arithmetic or learning the capitals of all the countries of the world, for example, are certainly activities that *indicate* intelligence in humans. The issue here is whether a computer system that can perform these tasks can be said to *know* or *understand* anything. This point has been discussed at length (see, e.g., Searle, 1980, and appended commentary) and will be avoided here by describing the *behaviors* themselves as intelligent, without commitment as to how to describe the machines that produce them.

Problem solving. The first big "successes" in AI were programs that could solve puzzles and play games. Techniques such as looking ahead several moves and dividing difficult problems into easier subproblems evolved, respectively, into the fundamental AI techniques of *search* and *problem reduction.* Today's programs play championship-level checkers and backgammon, as well as very good chess. Another problem-solving program, the one that does symbolic evaluation of mathematical functions, performs very well and is being used widely by scientists and engineers. Some programs can even improve their own performance with experience.

As discussed below, the open questions in this area involve abilities that human players exhibit but cannot articulate, such as the chess master's ability to see the board configuration in terms of meaningful patterns. Another basic open question involves the original conceptualization of a problem, called in AI the *choice of problem representation.* Humans often solve a problem by finding a way of thinking about it that makes the solution easy; AI programs, so far, must be told how to think about the problems they solve (i.e., the space in which to search for the solution).

Logical reasoning. Closely related to problem and puzzle solving was early work on logical deduction. Programs were developed that could "prove" assertions by manipulating a data base of facts, each represented by discrete data-structures just as they are represented by formulas in mathematical logic. These methods, unlike many other AI techniques, could be shown to be complete and consistent. That is, given a set of facts, the programs theoretically could prove all theorems that followed from the facts, and only those theorems. Logical reasoning has been one of the most persistently investigated subareas of AI research. Of particular interest are the problems of finding ways of focusing on only the relevant facts from a large data base and of keeping track of the justifications for beliefs and updating them when new information arrives.

Programming. Although perhaps not an obviously important aspect of human cognition, programming itself is an important area of research in AI. Work in this area, called *automatic programming,* has investigated systems that can write computer programs from a variety of descriptions of their purpose, such as examples of input/output pairs, high-level language descriptions, and even English-language descriptions of algorithms. Progress has been limited to a few, fully worked-out examples. Automatic-programming research may result not only in semiautomated systems for software development but also in AI programs that learn (i.e., modify their behavior) by modifying their own code. Related work in the theory of programs is fundamental to all AI research.

Language. The domain of language understanding was also investigated by early AI researchers and has consistently attracted interest. Programs have been written that retrieve information from a data base in response to questions posed in English, that translate sentences from one language to another, that follow instructions or paraphrase statements given in English, and that acquire knowledge by reading textual material and building an internal data base. Some programs have even achieved limited success in interpreting instructions that are spoken into a microphone rather than typed into the computer. Although these language systems are not nearly so good as people are at any of these tasks, they are adequate for some applications. Early successes with programs that answered simple queries and followed simple directions, and early failures at machine-translation attempts, have resulted in a sweeping change in the whole AI approach to language. The principal themes of current language-understanding research are the importance of :t amounts of *knowledge* about the subject being discussed and the role of *expectations*, based on the s :t matter and the conversational situation, in interpreting sentences. The state of the art of practical lang e programs is represented by useful "front ends" to a variety of software systems. These programs accept only in some restricted form; they cannot handle some of the nuances of English grammar and are usefu for interpreting sentences only within a relatively limited domain of discourse. Although there has been very limited success at translating AI results in language and speech-understanding programs into ideas about the nature of human language *processing*, the realization of the importance in language understanding of extensive background knowledge, and of the contextual setting and intentions of the speakers, has changed our notion of what language or a theory of language might be.

Learning. Certainly one of the most significant aspects of human intelligence is our ability to learn. However, this is an example of cognitive behavior that is so poorly understood that very little progress has been made in accomplishing it in AI systems. Although there have been several interesting attempts at this, including programs that learn from examples, from their own performance, or from advice from others, AI systems do not exhibit noticeable learning.

Robotics and vision. One area of AI research that is receiving increasing attention involves programs that manipulate robot devices. Research in this field has looked at everything from the optimal movement of robot arms to methods of planning a sequence of actions to achieve a robot's goals. Some robots "see" through a TV camera that transmits an array of information back to the computer. The processing of visual information is another very active, and very difficult, area of AI research. Programs have been developed that can recognize objects and shadows in visual scenes, and even identify small changes from one picture to the next, for example, for aerial reconnaissance. The true potential of this research, however, is that it deals with *artificial* intelligences in perceived and manipulable environments similar to our own.

Systems and languages. In addition to work directly aimed at achieving intelligence, the development of new tools has always been an important aspect of AI research. Some of the most important contributions of AI to the world of computing have been in the form of spin-offs. Computer-systems ideas like time-sharing, list processing, and interactive debugging were developed in the AI research environment. Specialized programming languages and systems, with features designed to facilitate deduction, robot manipulation, cognitive modeling, and so on, have often been rich sources of new ideas. Most recent among these has been the many knowledge-representation languages. These are computer languages for encoding knowledge as data structures and reasoning methods as procedures, developed over the last five years to explore a variety of ideas about how to build reasoning programs. Terry Winograd's 1979 article "Beyond Programming Languages" discusses some of his ideas about the future of computing, inspired in part by his research on AI.

Expert systems. Finally, the area of "expert," or "knowledge-based," systems has recently emerged as a likely area for useful applications of AI techniques (Feigenbaum, 1977). Typically, the user interacts with an expert system in a form of consultation dialogue, just as he (or she) would interact with a human expert in a particular area: explaining his problem, performing suggested tests, and asking questions about proposed solutions. Current experimental systems have performed very well in consultation tasks like chemical and geological data analysis, computer-system configuration, completion of income tax forms, and even medical diagnosis. Expert systems can be viewed as intermediaries between human experts, who interact with the systems in *knowledge-acquisition* mode, and human users, who interact with the systems in *consultation* mode. Furthermore, much research in this area of AI has focused on providing these systems with the ability to *explain* their reasoning, both to make the consultation more acceptable to the user and to help the human expert locate the cause of errors in the system's reasoning when they occur.

Because its imminent commercial applications are indicative of important changes in the field, much of the ensuing discussion of the role of AI in the study of mind will refer to the expert-systems research. That these systems

- "represent" vast amounts of knowledge obtained from *human* experts,
- are used as *tools* to solve difficult problems using this knowledge,
- can be viewed as *intermediaries* between human problem solvers,
- must *explain* their "thought processes" in terms that people can understand, and
- are worth a lot of *money* to people with real problems

are the essential points that will be true of all of AI someday, in fact, of computers in general, and will change the role that AI research plays in the scientific study of thought.

Open problems. Although there have been much activity and progress in the 25-year history of AI, some very central aspects of cognition have not yet been achieved by computer programs. Our abilities to reason about others' beliefs, to know the limits of our knowledge, to visualize things, to be "reminded" of relevant events, to learn, to reason by analogy, and to make plausible inferences, realize when they are wrong, and know how to recover from them are not at all understood.

It is a fact that these and many other *fundamental* cognitive capabilities may remain problematic for some time. But it is also a fact that computer programs have successfully achieved a level of performance on a range of "intelligent" behaviors unmatched by anything other than the human brain. AI's failure to provide some seemingly simple cognitive capabilities in computer programs becomes, in the view of AI to be presented in this paper, part of the set of phenomena to be explained by the new science.

AI and the Study of Mind

AI research in problem solving, language processing, and so forth has produced some impressive and useful computer systems. It has also influenced, and been influenced by, research in many other fields. What, then, is the relation between AI and the other disciplines that study the various aspects of mind, for example, psychology, linguistics, philosophy, and sociology?

AI certainly has a unique method—designing and testing computer programs—and a unique goal—making those programs seem intelligent. It has been argued from time to time that these attributes make AI independent of the other disciplines:

> Artificial Intelligence was an attempt to build intelligent machines without any prejudice toward making the system simple, biological, or humanoid. (Minsky, 1968, p. 7)

But one does not start from scratch in building the first program to accomplish some intelligent behavior; the ideas about how that program is to work must come from somewhere. Furthermore, most AI researchers *are* interested in understanding the human mind and actively seek hints about its nature in their experiments with their programs.

The interest within AI in the results and open problems of other disciplines has been fully reciprocated by interest in and application of AI research activity among researchers in other fields. Many experimental and theoretical insights in psychology and linguistics, at least, have been sparked by AI techniques and results. Furthermore, this flow is likely to increase dramatically in the future; its source is the variety of new phenomena displayed by AI systems—the number, quality, utility, and level of activity of which will soon dramatically increase. But first let us examine what kind of interactions have taken place between AI and the other disciplines.

The Language of Computation

As we defined it at the outset, AI is a branch of computer science. Its practitioners are trained in the various subfields of computer science: formal computing theory, algorithm design, hardware and operating-systems architecture, programming languages, and programming. The study of each of these subareas has produced a language of its own, indicating our understanding of the important known *phenomena* of computing. The underlying assumption of our research is that this language (which involves concepts like process, procedure, interpreter, bottom-up and top-down processing, object-oriented programming, and trigger) and the experience with computation that it embodies will, in turn, assist us in understanding the various *phenomena* of mind.

Before we go on to discuss the utility of these computational concepts, it should be stated that, in fact, our understanding of computation itself is quite limited. Von Neumann (1958) dreamed of an "information theory" of the nature of thinking:

> The body of experience which has grown up around the planning, evaluating, and coding of complicated logical and mathematical automata will be the focus of much of this information theory. . . . It would be very satisfactory if one could talk about a "theory" of such automata. Regrettably, what at this moment exists—and to what I must appeal—can as yet be described only as an imperfectly articulated and hardly formalized "body of experience." (p. 2)

And ten years later, in their superb treatise on perceptronlike automata, Minsky and Papert (1969) lament:

> We know shamefully little about our computers and their computations. . . . We know very little, for instance, about how much computation a job should require. . . . The immaturity shown by our inability to answer questions of this kind is exhibited even in the language used to formulate the questions. Word pairs such as "parallel" vs. "serial," "local" vs. "global," and "digital" vs. "analog" are used as if they referred to well-defined technical concepts. Even when this is true, the technical meaning varies from user to user and context to context. But usually they are treated so loosely that the species of computing machine defined by them belongs to mythology rather than science. (pp. 1-2)

There is still no adequate theory of computation for understanding the nature and scope of symbolic processes, but there is rapidly accumulating experience with computation of all sorts—useful new concepts emerge continually.

The Computational Metaphor

The discipline most closely related to AI is cognitive psychology. These two disciplines deal primarily with the same kinds of behaviors—perception, memory, problem solving. And they are siblings: Modern cognitive psychology emerged from its behavior-oriented precursors in conjunction with the rise of AI. That there *might* be a relation between the new field of AI and the traditional interests of psychologists was evident from the beginning:

> Our fundamental concern was to discover whether the cybernetic ideas have any relevance for psychology. The men who have pioneered in this area have been remarkably innocent of psychology. . . . There must be some way to phrase the new ideas so that they can contribute to and profit from the science of behavior that psychologists have created. (Miller, Galanter, and Pribram, 1960, p. 3)

What in fact happened was that the existence of computing served as an inspiration to traditional psychologists to begin to theorize in terms of internal, cognitive mechanisms. Use of the concepts of computation as metaphors for the processes of the mind strongly influenced the form of modern theories of cognitive psychology—for example, theories expressed in terms of memories and retrieval processes:

> Computers accept information, manipulate symbols, store items in "memory" and retrieve them again, classify inputs, recognize patterns, and so on. Whether they do these things just like people was less important than that they do them at all. The coming of the computer provided a much-needed reassurance that cognitive processes were real. (Neisser, 1976, p. 5)

The metaphorical use of the language of computation in describing mental processes was found to be, at least for a time, quite fertile ground for sprouting psychological theories.

> During a period of concept formation, we must be well aware of the metaphorical nature of our concepts. However, during a period in which the concepts can accommodate most of our questions about a given subject matter, we can afford to ignore their metaphorical origins and confuse our description of reality with that reality. (Arbib, 1972, p. 11)

When pioneering work by Newell, Shaw, and Simon and by other research groups showed that "programming up" their intuitions about how humans solve puzzles, find theorems, and so on was adequate to get impressive results, the link between the study of human problem-solving and AI research was firmly established.

Consider, for example, computer programs that play chess. Current programs are quite proficient—the best experimental systems play at the human "expert" level, but not as well as human chess "masters." The programs work by searching through a space of possible moves, that is, considering the alternative moves and their consequences several steps ahead in the game, just as human players do. These programs, even some of the earliest versions, could search through thousands of moves in the time it takes human players to consider only a dozen or so alternatives. The theory of optimal search, developed as a mathematical formalism (paralleling, as a matter of fact, much of the work on optimal decision theory in operations research) constitutes some of the core ideas of AI.

The reason that computers cannot beat the best human players is that looking ahead is not all there is to chess. Since there are too many possible moves to search exhaustively, even on the fastest imaginable computers, alternative moves (board positions) must be *evaluated* without knowing for sure which move will lead to a winning game, and this is one of those skills that human chess experts cannot make explicit. Psychological studies have shown that chess masters have learned to *see* thousands of meaningful configurations of pieces when they look at chess positions, which presumably helps them decide on the best move, but no one has yet suggested how to design a computer program that can identify these configurations.

For the lack of theory or intuitions about human perception and learning, AI progress on computer chess has virtually stopped, but it is quite possible that new insights into a very general problem were gained. The computer programs had pointed up, more clearly than ever, what would be useful for a cognitive system to learn to see. It takes many years for chess experts to develop their expertise—their ability to "understand" the game in terms of such concepts and patterns that they cannot explain easily, if at all. The general problem is of course, to determine what it is about our experience that we apply to future problem solving: What kind of *knowledge* do we glean from our experience? The work on chess indicated some of the demands that would be placed on this knowledge.

Language Translation and Linguistics

Ideas about getting computers to deal in some useful way with the human languages, called "natural" languages by computer scientists, were conceived before any machines were ever built. The first line of attack was to try to use large, bilingual dictionaries stored in the computers to translate sentences from one language to another (Barr and Feigenbaum, 1981, pp. 233-238). The machine would look up the translation of the words in the original sentence, figure out the "meaning" of the sentence (perhaps expressed in some *interlingua*), and produce a syntactically correct version in the target language.

It did not work. It became apparent early on that processing language in any useful way involved *understanding*, which in turn involved a great deal of knowledge about the world—in fact, it could be argued that the more one "knows," the more one "understands" each sentence one reads. And the level of world knowledge needed for any *useful* language-processing is much higher than our original intuitions led us to expect.

There has been a serious debate about whether AI work in computational linguistics has enlightened us at all about the nature of language (see Dresher and Hornstein, 1976, and the replies by Winograd, 1977, and Schank and Wilensky, 1977). The position taken by AI researchers is that if our goal in linguistics is to include understanding sentences like *Do you have the time?* and *We'll have dinner after the kids wash their hands*, which involve the total relationship between the speakers, then there is much more to it than the syntactic arrangement of words with well-defined meanings—that although the study in linguistics of the systematic regularities within and between natural languages is an important key to the nature of language and the workings of the mind, it is only a small part of the problem of building a *useful* language processor and, therefore, only a small part of an adequate understanding of language (Schank and Abelson, 1977):

> For both people and machines, each in their own way, there is a serious problem in common of making sense out of what they hear, see, or are told about the world. The conceptual apparatus necessary to perform even a partial feat of understanding is formidable and fascinating. (p. 2)

> Linguists have almost totally ignored the question of how human understanding works. . . . It has nevertheless been consistently regarded as important that computers deal well with natural language. . . . None of these high-sounding things are possible, of course, unless the computer really 'understands' the input. And that is the theoretical significance of these practical questions—to solve them requires no less than articulating the detailed nature of 'understanding'. If we understood how a human understands, then we might know how to make a computer understand, and vice versa. (p. 8)

This idea that building AI systems requires the articulation of the detailed nature of understanding, that is, that implementing a theory in a computer program requires one to "work out" one's fuzzy ideas and concepts, has been suggested as a major contribution of AI research (Schank and Abelson, 1977):

> Whenever an AI researcher feels he understands the process he is theorizing about in enough detail, he then begins to program it to find out where he was incomplete or wrong. . . . The time between the completion of the theory and the completion of the program that embodies the theory is usually extremely long. (p. 20)

And Newell (1970), in a thorough discussion of eight possible ways one might view the relation of AI to psychology, suggests that building programs "forces psychologists to become operational, that is, to avoid the fuzziness of using mentalistic terms" (p. 365).

Certainly the original conception of the machine-translation effort, although it was intuitively sensible, fell far short of what would be required to enable a machine to handle language, indicating a limited conception of what language is. It is in the broadening of this conception that AI has contributed most to the study of language (Schank and Abelson, 1977, p. 9). Thus, AI can show, as in the examples of chess and language understanding, that intuitive notions and assumptions about mental processes just do not *work*. Furthermore, analyzing the behavior of AI programs implemented on the basis of existing, inadequate concepts can offer hints on how the concepts of the theory affect the success of its application.

Scientific Languages and Theory Formation

Lawrence Miller, in a 1978 article that reviews the dialogue between psychologists and AI researchers about AI's contribution to the understanding of mind, concludes that

> the critics of AI believe that it is easy to construct plausible psychological theories; the difficult task is demonstrating that these theories are true. The advocates of AI believe that it is difficult to construct adequate psychological theories; but once such a theory has been constructed, it may be relatively simple to demonstrate that it is true. (p. 113)

And Schank and Abelson (1977) agree:

> We are not oriented toward finding out which pieces of our theory are quantifiable and testable in isolation. We feel that such questions can wait. First we need to know if we have a viable theory. (p. 21)

Just as AI must consider the same issues that psychology and linguistics address, other aspects of knowledge dealt with by other traditional disciplines must also be considered. For example, current ideas in AI about linking computing machines into coherent systems or cooperative problem-solvers forces us to consider the sociological aspects of knowing. A fundamental problem in AI is communication among many individual units, each of which "knows" some things relevant to some problems as well as something about the other units. The form of the communication between units, the organizational structure of the complex, and the nature of the individuals' knowledge of each other are all questions that must find some engineering solution if the apparent power of "distributed processing" is to be realized.

These issues have been studied in other disciplines, albeit from very different perspectives and with different goals and methods. We can view the different control schemes proposed for interprocess communication, for example, as attempts to design *social systems* of knowledgeable entities. Our intuitions, once again, form the specifications for the first systems. Reid G. Smith (1978) has proposed a *contract net* where the individual entities *negotiate* their roles in attacking the problem, via requests for assistance from other processors, proposals for help in reply, and contracts indicating agreement to delegate part of the problem to another processor; and Kornfeld and Hewitt (1981) have developed a model explicitly based on problem solving in the scientific community. Only after we have been able to build many systems based on such models will we be able to identify the key factors in the design of such systems.

There is another kind of study of the mind, conducted by scientists who seek to understand the workings of the brain. The brain as a mechanism has been associated with computing machines since their invention and has puzzled computer scientists greatly:

> We know the basic active organs of the nervous system (the nerve cells). There is every reason to believe that a very large-capacity memory is associated with this system. We do most emphatically *not* know what type of physical entities are the basic components for the memory in question. (von Neumann, 1958, p. 68)

If research on AI produces a language for describing what a computational system is doing, in terms of processes, memories, messages, and so forth, then that language may very well be the one in which the function of the neural mechanisms should be described (Lenat, 1981; Torda, 1982). And, as Herbert Simon (1980) points out, this functionality may be shared by nature's other brand of computing device, DNA:

> It might have been necessary a decade ago to argue for the commonality of the information processes that are employed by such disparate systems as computers and human nervous systems. The evidence for that commonality is now overwhelming, and the remaining questions about the boundaries of cognitive science have more to do with whether there also exist nontrivial commonalities with information processing in genetic systems than with whether men and machines both think. (p. 45)

One more example of the overlap of concerns between AI and the related disciplines is the following. Making it possible for an individual to know something about what another knows, without actually knowing it, involves defining the *nature* of what is known elsewhere: who the experts are on what kinds of problems and what they might know that could be useful. This relates directly to the categorization of knowledge that is the essence of library science. Instead of dealing with categories according to which static books will be filed, however, AI must consider the *dynamic* aspects of systems that know and learn.

The relation, then, between AI and disciplines like psychology, linguistics, sociology, brain science, and library science is a complex one. Certainly our current understanding of the phenomena dealt with by these disciplines—cognition, perception, memory, language, social systems, and categories of knowledge—has provided the intuitions and models on which the first AI programs were built. And, as has happened in psychology and linguistics, these first systems may, in turn, show us new aspects of the phenomena that we have not considered in studying their natural occurrence. But, most important, the development of AI systems, of *useful* computer tools for knowledge-oriented tasks, will expose us to many new phenomena and variations that will force us to increase our understanding.

The Practice of AI

AI, and computer science in general, employs a unique method among the disciplines involved in advancing our understanding of cognition—building computers and programs, and observing and trying to explain patterns in the behavior of these systems. The programs are the phenomena to be studied (Newell, 1981):

> Conceptual advances occur by (scientifically) uncontrolled experiments in our own style of computing. . . . The solution lies in more practice and more attention to what emerges there as pragmatically successful. (p. 4)

> Observing our own practice—that is, seeing what the computer implicitly tells us about the nature of intelligence as we struggle to synthesize intelligent systems—is a fundamental source of scientific knowledge for us. (p. 19)

Thus, AI is one of the "sciences of the artificial," as Herbert Simon (1969) has defined them in an influential paper. Half of the job is designing systems so that their performance will be interesting. There is a valuable heuristic in generating these designs: The systems that we are naturally inclined to want to build are those that will be *useful in our environment*. Our environment will shape them, as it shaped us. As Simon described the development of time-sharing systems:

> Most actual designs have turned out initially to exhibit serious deficiencies, and most predictions of performance have been startlingly inaccurate. Under these circumstances, the main route open to the development and improvement of time-sharing systems is to build them and see how they behave. (p. 21)

The Genus of Symbol Manipulators

Newell and Simon's psychologically phrased idea of "observing the behavior of programs" follows from their pioneering research program in what they have called information processing psychology. Newell and Simon developed, in the early years of this enterprise, some of the first computer programs that showed reasoning capabilities. This research on chess-playing, theorem-proving, and problem-solving programs was undertaken as an explicit attempt to model the corresponding human behaviors. But Newell and Simon took the strong position that these programs were not to serve simply as metaphors for human thought but were themselves theories. In fact, they argued that programs were the natural vehicle for expressing theories in psychology:

> An abstract concept of an information processing system has emerged with the development of the digital computers. In fact, a whole array of different abstract concepts has developed, as scientists have sought to capture the essence of the new technology in different ways. . . . With a model of an information processing system, it becomes meaningful to try to represent in some detail a particular man at work on a particular task. Such a representation is not metaphor, but a precise symbolic model on the basis of which pertinent specific aspects of the man's problem solving behavior can be calculated. (Newell and Simon, 1972, p. 5)

> Taking the view that artificial intelligence is theoretical psychology, simulation (the running of a program purporting to represent some human behavior) is simply the calculation of the consequences of a psychological theory. (Newell, 1973a, p. 47)

> A framework comprehensive enough to encourage and permit thinking is offered, so that not only answers, but questions, criteria of evidence, and relevance all become affected. (Newell, 1973a, p. 59)

Newell and Simon, in their view that computer programs are a vehicle for expressing psychological theories rather than just serving as a metaphor for mental processes, were already taking a strong position relative to even the new breed of cognitive psychologists who were talking in terms of computerlike mental mechanisms. As Paul R. Cohen (1982) puts it, in his review of AI work on models of cognition:

> We should note that we have presented the strongest version of the information-processing approach, that advocated by Newell and Simon. Their position is so strong that it defines information-processing psychology almost by exclusion: It is the field that uses methods alien to cognitive psychology to explore questions alien to AI. This is an exaggeration, but it serves to illustrate why there are thousands of cognitive psychologists, and hundreds of AI researchers, and very few information-processing psychologists. (p. 7)

However, Newell and Simon did not stop there. A further development in their thinking identified brains and computers as two species of the genus of *physical symbol systems*—the kind of system that, they argue, *must* underlie any intelligent behavior.

> At the root of intelligence are symbols, with their denotative power and their susceptibility to manipulation. And symbols can be manufactured of almost anything that can be arranged and patterned and combined. Intelligence is mind implemented by any patternable kind of matter. (Simon, 1980, p. 35)

> A physical symbol system has the necessary and sufficient means for general intelligent action. (Newell and Simon, 1976, p. 116)

> Information processing psychology is concerned essentially with whether a successful theory of human behavior can be found within the domain of symbolic systems. (Newell, 1970, p. 372)

> The basic point of view inhabiting our work has been that programmed computer and human problem solver are both species belonging to the genus IPS. (Newell and Simon, 1972, p. 869)

It is this view of computers—as systems that share a common, underlying structure with the human intelligence system—that promotes the behavioral view of AI computer research. Although these machines are not limited by the rules of development of their natural counterpart, they will be shaped in their development by the same natural *constraints* responsible for the form of intelligence in nature.

The Flight Metaphor

The question of whether machines could *think* was certainly an issue in the early days of AI research, although dismissed rather summarily by those who shaped the emerging science:

> To ask whether these computers can think is ambiguous. In the naive realistic sense of the term, it is people who think, and not either brains or machines. If, however, we permit ourselves the ellipsis of referring to the operation of the brain as "thinking," then, of course, our computers "think." (McCulloch, 1964, p. 368)

Addressing fundamental issues like this one in their early writing, several researchers suggested a parallel with the study of flight, considering cognition as another natural phenomenon that could eventually be achieved by machines:

> Today, despite our ignorance, we can point to that biological milestone, the thinking brain, in the same spirit as the scientists many hundreds of years ago pointed to the bird as a demonstration in nature that mechanisms heavier than air could fly. (Feigenbaum and Feldman, 1963, p. 8)

It is instructive to pursue this analogy a bit farther. Flight, as a way of dealing with the contingencies of the environment, takes many forms—from soaring eagles to hovering hummingbirds. If we start to study flight by examining its forms in nature, our initial understanding of what we are studying might involve terms like feathers, wings, weight-to-wing-size ratios, and probably wing-flapping, too. This is the *language* we begin to develop—identifying regularities and making distinctions among the phenomena. But when we start to build flying artifacts, our understanding changes immediately:

> Consider how people came to understand how birds fly. Certainly we observed birds. But mainly to recognize certain phenomena. Real understanding of *bird flight* came from understanding *flight;* not birds. (Papert, 1972, pp. 1-2)

Even if we fail a hundred times at building a machine that flies by flapping its wings, we learn from every attempt. And eventually we abandon some of the assumptions implicit in our definition of the phenomena under study and realize that flight does not require wing movement or even wings:

> Intelligent behavior on the part of a machine no more implies complete functional equivalence between machine and brain than flying by an airplane implies complete functional equivalence between plane and bird. (Armer, 1963, p. 392)

Every new design brings new data about what works and what does not, and clues as to why. Every new contraption tries some different *design alternative* in the space defined by our theory language. And every attempt clarifies our understanding of what it means to fly.

But there is more to the sciences of the artificial than defining the "true nature" of natural phenomena. The exploration of the artifacts themselves, the stiff-winged flying machines, because they are *useful* to society, will naturally extend the exploration of the various points of interface between the technology and society. While nature's exploration of the possibilities is limited by its mutation mechanism, human inventors will vary every parameter they can think of to produce effects that might be useful—exploring the constraints on the design of their machines from every angle. The space of "flight" phenomena will be populated by examples that nature has not had a chance to try.

Exploring the Space of Cognitive Phenomena

This argument, that the utility of intelligent machines will *drive* the exploration of their capabilities, suggests that the development of AI technology has begun an exploration of cognitive phenomena that will involve aspects of cognition that are not easy to study in nature. In fact, as with the study of flight, AI will allow us to see natural intelligence as a limited capability, in terms of the *design trade-offs* made in the evolution of biological cognition:

> Computer science is an empirical discipline. . . . Each new machine that is built is an experiment. . . . Each new program that is built is an experiment. It poses a question to nature, and its behavior offers clues to an answer. . . . We build computers and programs for many reasons. We build them to serve society and as tools for carrying out the economic tasks of society. But as basic scientists we build machines and programs as a way of discovering new phenomena and analyzing phenomena we already know about. . . . The phenomena surrounding computers are deep and obscure, requiring much experimentation to assess their nature. (Newell and Simon, 1976, p. 114)

For what will AI systems be useful? How will they be involved in the economic tasks of society? It has certainly been argued that this point is one that distinguishes biological systems from machines (Norman, 1980):

> The human is a physical symbol system, yes, with a component of pure cognition describable by mechanisms. . . . But the human is more: The human is an animate organism, with a biological basis and an evolutionary and cultural history. Moreover, the human is a social animal, interacting with others, with the environment, and with itself. The core disciplines of cognitive science have tended to ignore these aspects of behavior. (pp. 2-4)

> The difference between natural and artificial devices is not simply that they are constructed of different stuff; their basic functions differ. Humans survive. (p. 10)

Tools evolve and survive according to their utility to the people who use them. Either the users find better tools or their competitors find them. This process will certainly continue with the development of cognitive tools and will dramatically change the way we think about AI:

> We measure the intelligence of a system by its ability to achieve stated ends in the face of variations, difficulties and complexities posed by the task environment. This general investment of computer science in attaining intelligence . . . becomes more obvious as we extend computers to more global complex and knowledge-intensive tasks—as we attempt to make them our agents, capable of handling on their own the full contingencies of the natural world. (Newell and Simon, 1976, pp. 114-115)

In fact, this change has already begun in AI laboratories, but the place where the changing perception of AI systems is most dramatic and accelerated is, not surprisingly in our society, the marketplace.

AI, Inc.

To date, three of the emerging AI technologies have attracted interest as commercial possibilities: robots for manufacturing, natural-language front-ends for information-retrieval systems, and expert systems. The reason that a company like General Motors invests millions of dollars in robots for the assembly line is not scientific curiosity or propaganda about "retooling" their industry. GM believes these robots are essential to its economic survival. AI technology will surely change many aspects of American industry, but its application to real problems will just as surely change the emerging technology—change our perception of its nature and of its implications about knowledge. The remaining discussion will focus on this issue in the context of expert systems.

Expert Systems

With work on the DENDRAL system in the mid-1960s, AI researchers began pushing work on *problem-solving* systems beyond constrained domains like chess, robot planning, blocks-world manipulations, and puzzles: They started to consider symbolically expressed problems that were known to be difficult for the *best* human researchers to solve (see Lindsay, Buchanan, Feigenbaum, and Lederberg, 1980).

> One needs to move toward task environments of greater complexity and openness—to everyday reasoning, to scientific discovery, and so on. The tasks we tackled, though highly complex by prior psychological standards, still are simple in many respects. (Newell and Simon, 1972, p. 872)

Humans have difficulty keeping track of all of the knowledge that might be relevant to a problem, exploring all of the alternative solution-paths, and making sure none of the valid solutions is overlooked in the process. Work on DENDRAL showed that when human experts could explain exactly what they were doing in solving their problems, the machine could achieve expert-level performance.

Continued research at Stanford's Heuristic Programming Project next produced the MYCIN system, an experiment in modeling medical diagnostic reasoning (Shortliffe, 1976). In *production rules* of the form *If <condition> then <action>*, Shortliffe encoded the kind of information about the reasoning processes of physicians that they *were most able to give*—advice about what to do in certain situations. In other words, the *if* part of the rules contains clauses that attempt to differentiate a certain situation, and the *then* part describes what to do if one finds oneself in that situation. This production-rule *knowledge representation* worked surprisingly well: MYCIN was able to perform its task in a specific area of infectious-disease diagnosis as well as the best experts in the country.

Furthermore, the MYCIN structure was seen to be, at least to some extent, independent of the domain of medicine. So long as experts could describe their knowledge in terms of *If. . . then . . .* rules, the reasoning mechanism that MYCIN used to make inferences from a large set of rules would come up with the right questions and, eventually, a satisfactory analysis. MYCIN-like systems have been successfully built in research laboratories for applications as diverse as mineral exploration, diagnosis of computer-equipment failure, and even advising users about how to use complex systems.

Transfer of Expertise

There is an important shift in the view of expert systems just described that illustrates the changing perspective on AI that is likely to take place as it becomes an applied science. The early work on expert systems, building on AI research in problem solving, focused on representing and manipulating the facts in order to get answers. But through MYCIN, whose reasoning mechanism is actually quite shallow, it became clear that the way that these systems interacted with the people who had the knowledge and with those who needed it was an important, deep constraint on the system's architecture—on its knowledge representations and reasoning mechanisms:

> A key idea in our current approach to building expert systems is that these programs should not only be able to apply the corpus of expert knowledge to specific problems, but they should also be able to interact with the users and experts just as humans do when they learn, explain, and teach what they know. . . . These *transfer of expertise* (TOE) capabilities were originally necessitated by "human engineering" considerations—the people who build and use our systems needed a variety of "assistance" and "explanation" facilities. However, there is more to the idea of TOE than the implementation of needed user features: These social interactions—learning from experts, explaining one's reasoning, and teaching what one knows—are essential dimensions of human knowledge. They are as fundamental to the nature of intelligence as expert-level problem-solving, and they have changed our ideas about representation and about knowledge. (Barr, Bennett, and Clancey, 1979, p. 1)

Randall Davis's (1976) TEIRESIAS system, built within the MYCIN framework, was the first to focus on the *transferral* aspects of expert systems. TEIRESIAS offered aids for the experts who were entering knowledge into the system and for the system's users. For example, in order for an expert to figure out why a system has come up with the wrong diagnosis or is asking an inappropriate question, he (or she) has to understand its behavior in his own terms: The system must *explain* its reasoning in terms of concepts and procedures with which the expert is familiar. The same sort of explanation facility is necessary for the eventual user of an expert system who will want to be assured that the system's answers are well founded. Expert-systems technology had to be extended to facilitate such interactions, and, in the process, our conception of what an expert system was had changed. No longer did the systems simply solve problems; they now transferred expertise from people who had it to people who could use it:

> We are building systems that take part in the human activity of *transfer of expertise* among experts, practitioners, and students in different kinds of domains. Our problems remain the same as they were before: We must find good ways to represent knowledge and meta-knowledge, to carry on a dialogue, and to solve problems in the domain. But the guiding principles of our approach and the underlying constraints on our solutions have subtly shifted: Our systems are no longer being designed solely to be expert problem solvers, using vast amounts of encoded knowledge. There are aspects of "knowing" that have so far remained unexplored in AI research: By participation in *human* transfer of expertise, these systems will involve more of the fabric of behavior that is the reason we *ascribe* knowledge and intelligence to people. (Barr, Bennett, and Clancey, 1979, p. 5)

The Technological Niche

It is the goal of those who are involved in the commercial development of expert-systems technology to incorporate that technology into some device that can be sold. But the *environment* in which expert systems operate is our own cognitive environment; it is within this sphere of activity—people solving their problems—that the eventual expert-system products must be found useful. *They will be engineered to our minds.*

> With these systems, it will at last become economical to match human beings in real time with really large machines. This means that we can work toward programming what will be, in effect, "thinking aids." In the years to come we expect that these man-machine systems will share, and perhaps for a time be dominant, in our advance toward the development of "artificial intelligence." (Minsky, 1963, p. 450)

It is a long way from the expert systems developed in the research laboratories to any products that fit into people's lives; in fact, it is difficult even to envision what such products will be. Egon Loebner of Hewlett-Packard Laboratories tells of a conversation he had many years ago with Vladimir Zworykin, the inventor of television technology. Loebner asked Zworykin what he had in mind for his invention when he was developing the technology in the 1920s—what kind of product he thought his efforts would produce. The inventor said that he had had a very clear idea of the eventual use of TV: He envisioned medical students in the gallery of an operating room getting a clear picture on their TV screens of the details of the operation being conducted below them.

One cannot, at the outset, understand the application of a new technology, because it will find its way into realms of application that do not yet exist. Loebner has described this process in terms of the *technological niche*, paralleling modern evolution theory (Loebner, 1976; Loebner and Borden, 1969). Like the species and their environment, inventions and their applications are co-defined—they constantly evolve together, with niches representing periods of relative stability, into a new reality:

> Moreover, the niches themselves are . . . defined in considerable measure by the whole constellation of organisms themselves. There can be no lice without hairy heads for them to inhabit, nor animals without plants. (Simon, 1980, p. 44)

Thus, technological inventions change as they are applied to people's needs, and the activities that people undertake change with the availability of new technologies. And as people in industry try to push the new technology toward some profitable niche, they will also explore the nature of the underlying phenomena. Of course, it is not just the scientists and engineers who developed the new technology who are involved in this exploration: Half of the job involves finding out what the new capabilities can do for people.

Recognition of the commercial application of TV technology was accomplished by David Sarnoff, after the model he had used for the radio broadcasting industry. It is important to note that the "commercial product" that resulted from TV technology, the TV-set receiver, was only part of a gigantic *system* that had to be developed for its support (actually imported from radio, with modifications and extensions), involving broadcast technology, the networks, regulation of the air waves, advertising, and so forth. Loebner refers to this need for *systemwide* concern with product development as the Edisonian model of technological innovation: Edison's achievement of the invention of the long-life, commercially feasible light bulb was conducted in parallel with his successful development of the first dynamo for commercially producing electric power and with his design and implementation of the first electric-power distribution network.

The Knowledge Industry

Among the scientific disciplines that study knowledge, the potential for commercial applications of artificial intelligence presents unique opportunities. To identify and fill the niches in which intelligent machines will survive, we must ask questions about "knowledge" from a rather different perspective. We must identify the role that the various aspects of intelligence play, or could play, in the affairs of men, in such a way that we can identify correctable shortcomings in how things are done.

There is no question that the current best design of an intelligent system, the human brain, has its limitations. Computers have already helped people deal with such shortcomings as memory failure and confusions, overloading in busy situations, their tendency to boredom, and their need for sleep. These extended capabilities—total recall, rapid processing, and uninterrupted attention—are cognitive capabilities that we have been willing to concede to the new species in the genus of *symbol manipulators*. They have helped us do the things we did before, and have made some entirely new capabilities possible, for example, airline reservation systems, 24-hour banking, and Pac-Man (although the truly challenging computer "games" are yet to come!). Intelligence is also going to be present in this new species, as envisioned 20 years ago by Marvin Minsky (1963):

> I believe . . . that we are on the threshold of an era that will be strongly influenced, and quite possibly dominated, by intelligent problem-solving machines. (p. 406)

Finding a way to apply this new intellectual capability, for effectively applying relevant experience to new situations, is the task ahead for AI, Inc.

> We have hardly begun to understand what this abundant and cheap intellectual power will do to our lives. It has already started to change physically the research laboratories and the manufacturing plants. It is difficult for the mind to grasp the ultimate consequences for man and society. (Riboud, 1979)

It may be a while in coming, and it may involve a rethinking of the way we go about some cognitive activities. But it is extremely important that the development of intelligent machines be pursued, for the human mind not only is limited in its storage and processing capacity but it also has known bugs: It is easily misled, stubborn, and even blind to the truth, especially when pushed to its limits.

And, as is nature's way, everything gets pushed to the limit, including humans. We must find a way of organizing ourselves more effectively, of bringing together the energies of larger groups of people toward a common goal. Intelligent systems, built from computer and communications technology, will someday know more than any individual human about what is going on in complex enterprises involving millions of people, such as a multinational corporation or a city. And they will be able to explain each person's part of the task. We will build more productive factories this way, and maybe someday a more peaceful world. We must keep in mind, following our analogy of flight, that the capabilities of intelligence as it exists in nature are not necessarily its natural limits:

> There are other facets to this analogy with flight; it, too, is a continuum, and some once thought that the speed of sound represented a boundary beyond which flight was impossible. (Armer, 1963, p. 398)

Bibliography

Arbib, M. A. 1972. *The metaphorical brain.* New York: Wiley-Interscience.

Armer, P. 1963. Attitudes toward intelligent machines. In E. A. Feigenbaum and J. Feldman (Eds.), *Computers and thought.* New York: McGraw-Hill, 389-405.

Barr, A., Bennett, J. S., and Clancey, W. J. 1979. *Transfer of expertise: A theme for AI research* (Working Paper No. HPP-79-11). Stanford University, Heuristic Programming Project.

Barr, A., and Feigenbaum, E. A. (Eds.). 1981. *The handbook of artificial intelligence* (Vol. 1). Los Altos, Calif.: Kaufmann.

Becker, J. D. 1975. Reflections on the formal description of behavior. In D. G. Bobrow and A. Collins (Eds.), *Representation and understanding: Studies in cognitive science.* New York: Academic Press, 83-102.

Bernstein, J. 1981. Profiles: Marvin Minsky. *New Yorker*, December 14, pp. 50-126.

Cohen, P. R. 1982. Models of cognition: Overview. In P. R. Cohen and E. A. Feigenbaum (Eds.), *The handbook of artificial intelligence* (Vol. 3). Los Altos, Calif.: Kaufmann, 1-10.

Davis, R. 1976. *Applications of meta-level knowledge to the construction, maintenance, and use of large knowledge bases* (Tech. Rep. STAN-CS-76-564). Stanford University, Computer Science Department. (Reprinted in R. Davis and D. Lenat (Eds.), *Knowledge-based systems in artificial intelligence.* New York: McGraw-Hill, 1982, 229-490.)

Dresher, B. E., and Hornstein, N. 1976. On some supposed contributions of artificial intelligence to the scientific study of language. *Cognition* 4(4):321-398. (See also their replies to Schank and Wilensky, *Cognition* 5:147-150, and to Winograd, *Cognition* 5:379-372.)

Feigenbaum, E. A. 1977. The art of artificial intelligence, I: Themes and case studies of knowledge engineering. *Proceedings of the Fifth International Joint Conferences on Artificial Intelligence*, 1014-1029.

Feigenbaum, E. A., and Feldman, J. (Eds.). 1963. *Computers and thought.* New York: McGraw-Hill.

Kornfeld, W. A., and Hewitt, C. 1981. *The scientific community metaphor* (Tech. Rep. AIM-641). Massachusetts Institute of Technology, AI Laboratory.

Lenat, D. G. 1981. *The heuristics of nature* (Working Paper No. HPP-81-22). Stanford University, Heuristic Programming Project.

Lindsay, R., Buchanan, B. G., Feigenbaum, E. A., and Lederberg, J. 1980. *DENDRAL.* New York: McGraw-Hill.

Loebner, E. E. 1976. Subhistories of the light emitting diode. *IEEE Transactions on Electron Devices* 23(7):675-699.

Loebner, E. E., and Borden, H. 1969. Ecological niches for optoelectronic devices. *WESCON*, Vol. 13, Session 20, 1-8.

Marr, D. 1977. Artificial intelligence—A personal view. *Artificial Intelligence* 9(1):1-13.

Maturana, H. 1976. Biology of language: The epistemology of reality. In *Psychology and biology of language and thought*. Ithaca, N.Y.: Cornell University Press.

McCorduck, P. 1979. *Machines who think*. San Francisco: Freeman.

McCulloch, W. 1964. The postulational foundations of experimental epistemology. In *Embodiments of mind*. Cambridge, Mass.: MIT Press, 359-372.

Miller, G. A., Galanter, E., and Pribram, K. H. 1960. *Plans and the structure of behavior*. New York: Holt, Rinehart and Winston.

Miller, L. 1978. Has artificial intelligence contributed to an understanding of the human mind? A critique of arguments for and against. *Cognitive Science* 2(2):111-128.

Minsky, M. 1963. Steps toward artificial intelligence. In E. A. Feigenbaum and J. Feldman (Eds.), *Computers and thought*. New York: McGraw-Hill, 406-450.

Minsky, M. (Ed.). 1968. *Semantic information processing*. Cambridge, Mass.: MIT Press.

Minsky, M., and Papert, S. 1969. *Perceptrons: An introduction to computational geometry*. Cambridge, Mass.: MIT Press.

Neisser, U. 1976. *Cognition and reality*. San Francisco: Freeman.

Newell, A. 1970. Remarks on the relationship between artificial intelligence and cognitive psychology. In R. Banerji and M. D. Mesarovic (Eds.), *Theoretical approaches to non-numerical problem solving*. New York: Springer-Verlag, 363-400.

Newell, A. 1973a. Artificial intelligence and the concept of mind. In R. Schank and K. Colby (Eds.), *Computer models of thought and language*. San Francisco: Freeman, 1-60.

Newell, A. 1973b. You can't play 20 questions with nature and win. In W. G. Chase (Ed.), *Visual information processing*. New York: Academic Press, 283-308.

Newell, A. 1980. Physical symbol systems. *Cognitive Science* 4(2):135-183.

Newell, A. 1981. The knowledge level. *AI Magazine* 2(2):1-20.

Newell, A., and Simon, H. A. 1972. *Human problem solving*. Englewood Cliffs, N.J.: Prentice-Hall.

Newell, A., and Simon, H. A. 1976. Computer science as empirical inquiry: Symbols and search (Turing Award Lecture, Association for Computing Machinery). *Communications of the ACM* 19(3):113-126.

Nilsson, N. 1974. Artificial intelligence. In J. L. Rosenfeld (Ed.), *Proceedings of the IFIP Congress* (Vol. 4). New York: American Elsevier, 778-801.

Nilsson, N. 1980. *Principles of artificial intelligence*. Palo Alto, Calif.: Tioga Press.

Norman, D. A. 1980. Twelve issues for cognitive science. *Cognitive Science* 4(1):1-32.

Papert, S. 1972. Paper given at the NUFFIC summer course on process models in psychology. The Hague: NUFFIC.

Riboud, J. 1979. Address to the meeting of shareholders, Schlumberger Limited.

Schank, R., and Abelson, R. 1977. *Scripts, plans, goals, and understanding.* Hillsdale, N.J.: Erlbaum.

Schank, R., and Wilensky, R. 1977. Response to Dresher and Hornstein. *Cognition* 5:133-146.

Searle, J. R. 1980. Minds, brains, and programs. *Behavioral and Brain Sciences* 3(3):411-457.

Shortliffe, E. H. 1976. *Computer-based medical consultations: MYCIN.* New York: American Elsevier.

Simon, H. A. 1969. *The sciences of the artificial.* Cambridge, Mass.: MIT Press.

Simon, H. A. 1980. Cognitive science: The newest science of the artificial. *Cognitive Science* 4(1):33-46.

Smith, R. G. 1978. *A framework for problem solving in a distributed processing environment* (Tech. Rep. STAN-CS-78-7009). Stanford University, Department of Computer Science. (Doctoral dissertation.)

Torda, C. 1982. *Information processing by the central nervous system and the computer (a comparison).* Berkeley, Calif.: Walters.

Turing, A. M. 1950. Computing machinery and intelligence. *Mind* 59:433-460.

von Neumann, J. 1958. *The computer and the brain.* New Haven, Conn.: Yale University Press.

Winograd, T. 1977. On some contested suppositions of generative linguistics about the scientific study of language. *Cognition* 5:151-179.

Winograd, T. 1979. Beyond programming languages. *Communications of the ACM* 22(7):391-401.

DATE
ILMED

- 8